Phenotree: Using Patterns of Independent Loss to Link Traits and Conserved Noncoding Elements

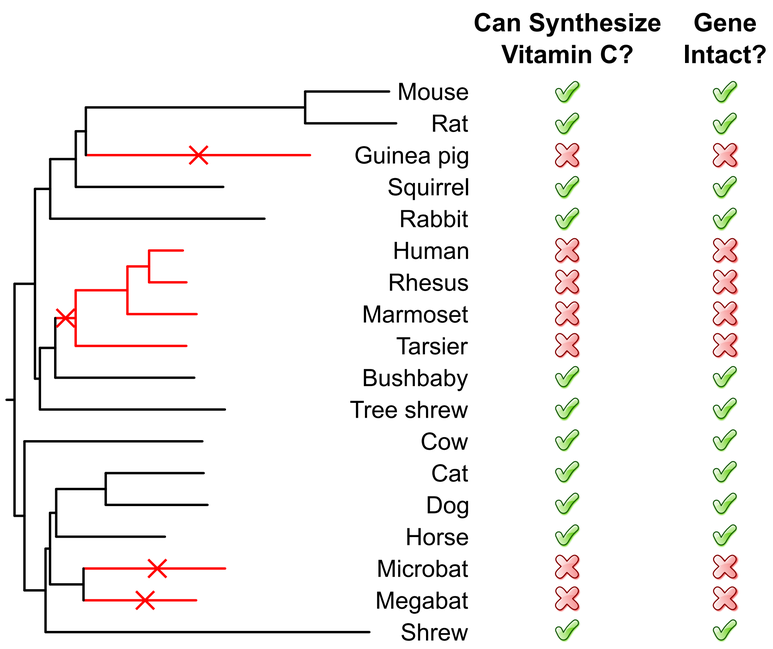
Robin Jia[1], Gray Camp[2], and Gill Bejerano[1,2]
Departments of 1. Computer Science and 2. Developmental Biology, Stanford University

Introduction

Many regions of the genome are evolutionarily conserved to some species but are not present in others. Such a pattern may arise if during evolution, one set of species acquired inactivating mutations in the same previously functional genomic region. Following such an event, there would be no pressure from natural selection to preserve the region, so it would likely accumulate random mutations and decay. With computational methods, we can detect some of these shifts from purifying to neutral selection.

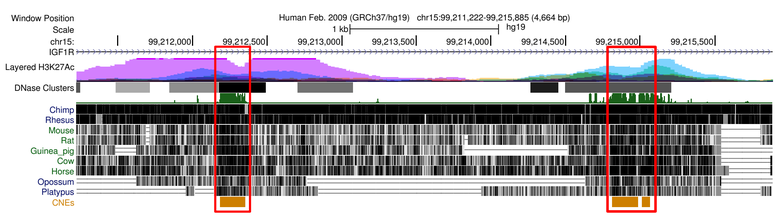
Moreover, we can use evolutionary history to infer the function of genomic regions. In this project, we searched for conserved elements that have been independently lost in multiple lineages. We simultaneously obtained vectors of traits scored over many mammalian species. Then, we looked for pairs of elements and traits that had the same pattern of loss. Since the probability of such a match occurring by chance is low, we hypothesize that at least some of these pairs are biologically related. As a proof of concept, a similar screen was able to trace the inability of some species to synthesize Vitamin C to the loss of the gene Gulo[1].

Independent Loss of Vitamin C Synthesis



Conserved Noncoding Elements

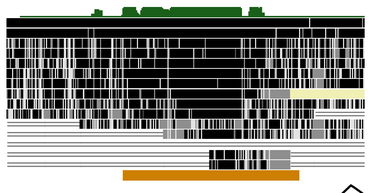
Many conserved regions of the genome do not code for proteins, yet their evolutionary constraint suggests that they have important biological functions. While some of these conserved noncoding elements (CNEs) are transcribed into functional noncoding RNAs, many may function as cis-regulatory elements, influencing the transcription of nearby genes via transcription factor binding. We focused on CNE loss in this screen, and hope to identify conserved regulatory elements responsible for biologically interesting phenotypes.



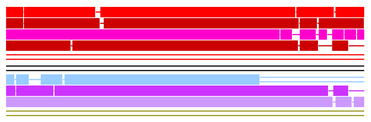
Methods

Percent Identity Computation

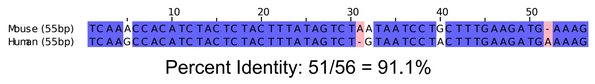
330,470 human CNEs



UCSC Chains from 39 other mammalian species

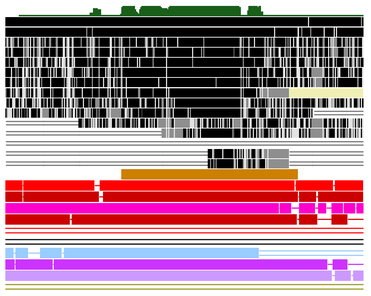


Local Pairwise Alignments

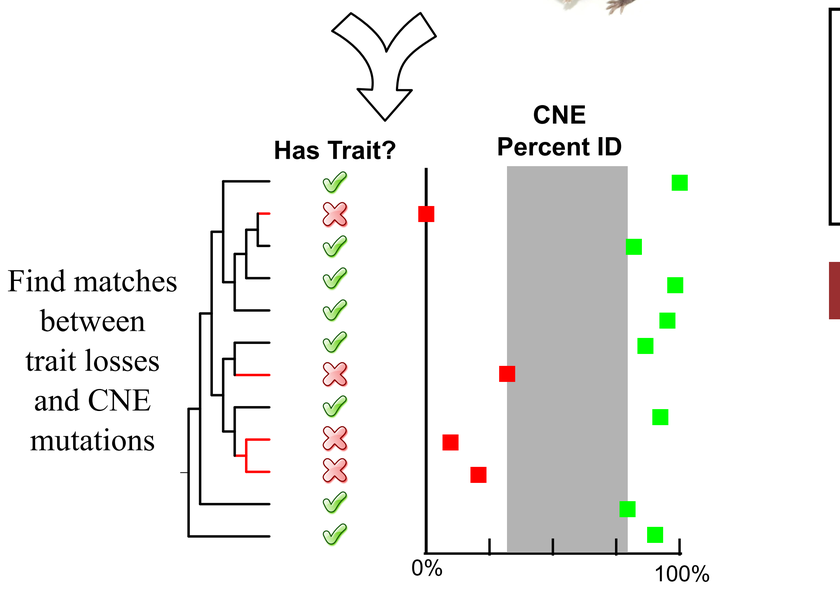


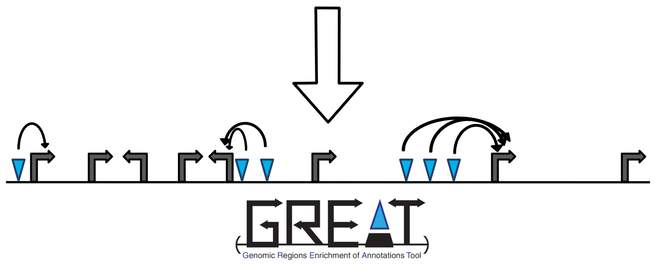
Matching CNEs with Trait Patterns

330,470 human CNEs with percent ID computed with 39 other mammals



4,500 traits scored for 22 of these mammalian species





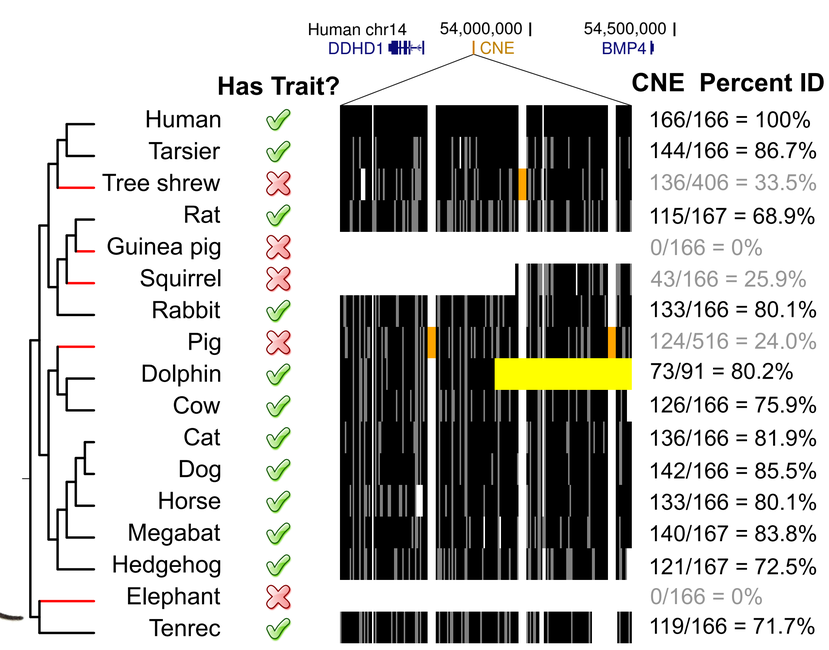
Associate CNEs with nearby genes via GREAT[2]

PubMed

Dispatch automated queries to PubMed to search for supporting literature

Results

Our screen uncovered numerous CNE-trait pairs that appear to be biologically interesting. For example, the screen associated the presence of an exposed region of the temporal bone near the ear with presence of a CNE in the GREAT regulatory domain of the gene BMP4 (bone morphogenic protein 4). BMP4 is widely known to be important in bone formation, and a coding SNP in BMP4 has been associated with otosclerosis, a form of hearing loss caused by excessive bone growth[3].



Future Directions

- Investigate how the relevance of hits changes as we vary the number of independent loss events required and the minimum percent identity difference separating species with and without the trait.
- Use CNEs from mouse to capture human losses
- Develop metrics to capture the asymmetry between "loss" species with high percent ID and "preserving" species with low percent ID
- Consider alternatives to percent identity for identifying lost CNEs (e.g. deletions, TF binding site affinities)
- Investigate alternatives to parsimony, such as probabilisitc models that use branch lengths
- Generate large sets of CNEs per trait and look for statistical enrichment of elements around relevant genes.
- Test candidates experimentally by (i) performing a reporter assay to determine if the lost CNE controls gene expression relevant to the corresponding trait, and (ii) mutating the element in a mouse model (if present) to measure the phenotypic consequences of CNE loss.

References: 1. Hiller et al., "A 'forward genomics' approach links genotype to phenotype using independent phenotypic losses among related species," *Cell Reports*, 2012. 2. McLean et al., "GREAT improves functional interpretation of cis-regulatory regions," *Nature Biotech.*, 2010. 3. Schrauwen et al., "Association of bone morphogenetic proteins with otosclerosis," *J. Bone Mineral Research*, 2008.